

WHAT IS CLAIMED IS:

1. A computer system for generating standardized product data, the computer system comprising:

a database operable to maintain data for a plurality of known products, each known product associated with a respective standardized product code; and

a processing facility coupled to the database, the processing facility operable to receive raw data for an unidentified product from a plurality of diverse data sources each of which has its own separate identifier for the unidentified product, to compare the raw data for the unidentified product against the data for the plurality of known products, and if there is a match between the raw data for the unidentified product and the data for one of the plurality of known products, to assign the respective standardized product code of the matching known product to the unidentified product.

2. The computer system of Claim 1 wherein:

the raw data comprises a raw description for the unidentified product;

the data maintained in the database comprises a separate stored description for each of the plurality of known products; and

the processing facility is operable to compare the raw description for the unidentified product against the stored descriptions for each of the plurality of known products.

3. The computer system of Claim 1 wherein:
the raw data comprises a number of field values for the
unidentified product;

5 the data maintained in the database comprises separate field
values for each of the plurality of known products; and

the processing facility is to compare a predetermined
combination of the field values for the unidentified product
against corresponding field values for each of the plurality of
10 known products.

4. The computer system of Claim 1 wherein the processing
facility is operable to parse the raw data into a number of
separate fields values for the unidentified product.

15 5. The computer system of Claim 1 wherein the processing
facility is operable to generate at least one guess as to a known
product which is a possible match for the unidentified product.

20 6. The computer system of Claim 5 wherein the processing
facility is operable to generate a confidence measure for the at
least one guess.

25 7. The computer system of Claim 1 further comprising an
interface coupled to the processing facility, the interface

operable to present the assigned standardized product code to an analyst for auditing.

8. A method performed on a computer system for generating
5 standardized product data, the method comprising:

maintaining data for a plurality of known products, each known product associated with a respective standardized product code;

receiving raw data for an unidentified product from a
10 plurality of diverse data sources, each data source having its own separate identifier for the unidentified product;

comparing the raw data for the unidentified product against the data for the plurality of known products; and

if there is a match between the raw data for the
15 unidentified product and the data for one of the plurality of known products, assigning the respective standardized product code of the matching known product to the unidentified product.

9. The method of Claim 8 comprising presenting the
20 assigned standardized product code to an analyst for auditing.

10. The method of Claim 8 comprising parsing the raw data into a number of separate fields values for the unidentified product.

11. The method of Claim 8 wherein the raw data comprises a raw description for the unidentified product and the maintained data comprises a separate stored description for each of the plurality of known products, and wherein comparing comprises
5 comparing the raw description for the unidentified product against the stored descriptions for each of the plurality of known products.

12. The method of Claim 8 wherein the raw data comprises a
10 number of field values for the unidentified product and the maintained data comprises separate field values for each of the plurality of known products, and wherein comparing comprises comparing a predetermined combination of the field values for the unidentified product against corresponding field values for each
15 of the plurality of known products.

13. The method of Claim 8 comprising generating at least one guess as to a known product which is a possible match for the unidentified product.
20

14. The method of Claim 13 comprising presenting the at least one guess to an analyst for assigning a standardized product code to the unidentified product.

15. The method of Claim 13 wherein the raw data comprises a
25 number of field values for the unidentified product and the

maintained data comprises separate field values for each of the plurality of known products, and wherein generating comprises performing a pattern comparison of the field values for the unidentified product against the field values for each known product.

16. A computer system for generating standardized product data, the computer system comprising:

a database operable to maintain data for a plurality of known products, each known product associated with a respective standardized product code, the data maintained in the database comprising a separate stored description and set of field values for each of the plurality of known products;

a processing facility coupled to the database and operable to:

receive raw data for an unidentified product from a plurality of diverse data sources each of which has its own separate identifier for the unidentified product, the raw data comprising a raw description and set of field values for the unidentified product,

compare the raw description for the unidentified product against the stored descriptions for each of the plurality of known products,

if the raw description for the unidentified product does not match any of the stored descriptions for the plurality of known products, compare a predetermined

combination of the field values for the unidentified product against corresponding field values for each of the plurality of known products, and

5 if the raw description for the unidentified product matches a stored description for one of the plurality of known products, or if all of the field values for the unidentified product match the corresponding field values for one of the plurality of known products for the predetermined combination, assign the respective
10 standardized product code of the matching known product to the unidentified product.

17. The computer system of Claim 16 wherein the processing facility is operable to parse the raw data into the fields values
15 for the unidentified product.

18. The computer system of Claim 16 comprising an interface coupled to the processing facility, the interface operable to present the assigned standardized product code to an analyst for
20 auditing.

19. The computer system of Claim 16 wherein the processing facility is operable to generate at least one guess as to a known product which is a possible match for the unidentified product.

25

20. The computer system of Claim 19 wherein the processing facility is operable to generate a confidence measure for the at least one guess.

5 21. The computer system of Claim 20 comprising an interface coupled to the processing facility, the interface operable to present the at least one guess and confidence measure to an analyst for assignment of a standardized product code.

10 22. A method performed on a computer system for generating standardized product data, the method comprising:

receiving raw data for an unidentified product from a plurality of diverse data sources each of which has its own separate identifier for the unidentified product, the raw data comprising a raw description and set of field values for the unidentified product;

comparing the raw description for the unidentified product against the stored descriptions for each of the plurality of known products;

20 if the raw description for the unidentified product does not match any stored description, comparing a predetermined combination of the field values for the unidentified product against corresponding field values for each of the plurality of known products; and

25 if the raw description for the unidentified product matches a stored description for one of the plurality of known products,

